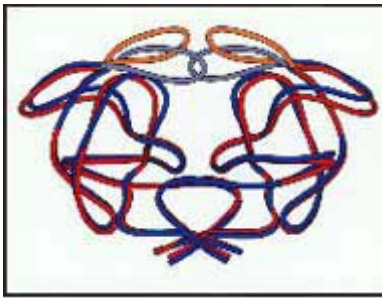
Examination of Large-Scale Motion in HIV-1 Protease

A Research Project of Dr. Steven W. Rick

Proteins often undergo large conformational changes as part of their function. These conformational changes may be the result of ligand binding in order to position the substrates correctly for catalysis, or the changes may be part of a receptor-signaling pathway. Structural changes in other proteins have important consequences in Alzheimer's disease and prion diseases. On the other hand, observed conformational differences in X-ray structures may be the results of crystallization conditions or crystal lattice forces, and are not necessarily related to the protein's function. Computer simulations can help to characterize the energetic and dynamic relationships of the different conformations; however, the time scales of these changes for proteins are typically too long for conventional molecular dynamics or Monte Carlo simulations. These problems preset a challenge both to computational resources and theoretical methods, requiring specialized techniques. In addition to studying the energetic relationships between two known structures or proteins, we are also interested in predicting how the ligand-enzyme structure changes as a result of mutations.

Alternate Conformations of the Same Protein

One protein which has demonstrated two stable conformations is the protease of the human immunodeficiency virus, type 1 (HIV-1). Inhibitors of HIV-1 protease are potent therapeutics for the acquired immune deficiency syndrome (AIDS) and, in the pursuit of inhibitors, many crystal structures of the enzyme have been determined, perhaps more than for any other protein. There structures indicate a significant difference between the ligand-free and inhibitor-bound enzyme. The enzyme is a homodimer with each subunit containing 99 residues. The active site region is capped by two identical ß-hairpin loops (the flaps) from each monomer. In the inhibitor-bound structures the flaps adopt a closed conformation, forming direct contacts with the substrate. In the ligand-free structures the flaps are more open with the position of the tip residues about 7Å from the inhibitor-bound conformation. The flaps are therefore flexible and must move in order for substrates to bind since, in all of the known structures, they block access to the active site. In addition, HIV-1 protease and the closely homologous protease of the simian immunodeficiency virus (SIV) have been observed to adopt the closed conformation in the absence of a ligand. The open and the closed forms have root-mean-square overlap of Ca atoms of 2 Å, and the differences between them cannot be represented in terms of a few degrees of freedom, such as a dihedral angle. One general method to calculate the difference in stability between the two very different structures is to find a series of intermediate structures which represent an incremental conversion from one structure to the other. The free energy difference between two similar structures can be readily calculated and the entire free energy difference between the initial and final structure can be pieced together from the free energy change between each neighbor. This involves calculating a path between the two structures represented by the series of replicas. Finding a minimum energy path is done by minimizing the energy of each structure, but in such a way that keeps the path straight and the structures evenly spaced along the path. The minimization is with respect to all atoms of the system (HIV protease has about 3000 atoms) for each replica (this application required about 100 replicas). THe calculations needed to compute the free energy differences are even larger. Once the path is found, then the free energy along the path can be calculated by performing a series of simulations, one for each replica point along the path. These calculations, requiring enough time to adequately sample the phase space around each point (about 50 ps) and also the addition of 3000 water molecules to represent the solvent environment, take about 2 months to complete running around the clock on a fast Digital ALPHA computer. For ligand-free HIV-1 protease, these calculations predict that the free energy of the open structure is less than the closed structures by -7±3 kcal/mol. The open form is the structure observed by X-ray crystallography. Interestingly, the more stable structure does not appear to have a lower potential energy, but rather is stabilized by entropy. The entropy for the transition, -TDS, is -11±10 kcal/mol and the energy change is 4±12 kcal/mol. The increased conformational freedom of the flap tips in the open structure is evident from X-ray crystallography experiments as well. For the open structures, the B-factors which relate to the dynamic motion of the Ca atoms of the flap tips (residues 49-51) are roughly twice as large as the average for all the Ca atoms. For the closed structures, including those with no substrate, the B-factors of the flap Ca atoms are not different from the average. The closed structure is stabilized by stronger interactions between the flaps, most importantly a hydrogen bond between the two flap tips. The transition occurs through a collapse and reformation of the ß structure of the conformationally flexible, glycine-rich flap tips rather than moving as rigid domains.



F-1. C-alpha traces of the open (red) and closed (blue) structures of HIV-1 protease, with flap residues 45-56 of each monomer shown in gray (closed) and orange (open).

Mutations and Conformational Relaxation

Inhibitors of HIV-1 protease are potent therapies for AIDS, however, mutants resistant to protease inhibitors can emerge in vivo in a short time. The in vivo mutation patterns are consistent with the in vitro data for the effects of the mutations on the binding constant, Ki. One common mutation which arises against a number of inhibitors is a change of isoleucine 84 to valine (I84V). Free energies for the change in Ki due to the mutation can be calculated using molecular dynamics simulations be free

energy perturbation or thermodynamic integration techniques. These methods do not depend on empirical parameterization o free energy terms and are exact calculations of the free energies, depending only on the quality of the interaction potentials used. Similar to the conformational calculations described above, the thermodynamic integration calculations are done by proceeding along a path (in this case an unphysical one) in which the residues are slowly converted from one amino acid to another, which once again requires large amounts of computer time. In addition to predicting free energy changes associated with the mutations, these calculations can also increase our understanding of the molecular origins of resistance and may hel in the development of better anti-viral strategies. The effects of this mutation on the binding constants, Ki, of the three inhibitors, L-735,542 (indinavir or MK-639), Ro 31-8959 (saquinavir), and KNI-272, to HIV type I (HIV-1) protease have been calculated using a thermodynamic integration method. Mutations which lead to resistance can occur in many locations on the enzyme; I84 is one of the residues with side chains in direct contact with the inhibitor - in the binding pockets S1, S1', S2 and S2'. The Cd methyl group of the Ile, which will be absent in the mutation to a valine, is in the S1 and S1' pockets. HIV protease is a homodimer, each monomer containing 99 residues. (The residues of the second monomer are denoted with the prime superscript). The calculated free energy changes for the three different inhibitors are all within 1 kcal/mole of the experimental results and the trend is also in agreement. The free energy change due to the mutation of I84', rather than I84, changes the most among the three inhibitors. The size of this term is consistent with the bulkiness of the group in this pocket and so the mutation of the Ile to a smaller Val would be expected to leave the largest cavity between the enzyme and the inhibitor. Ro 31- 8959 has the largest group in this pocket. An analysis of the trajectories generated in the free energy calculations reveals that cavities are created by the mutation. The size of the cavities induced by the mutation correlates well with the free energy changes, with a coefficient relating the cavity volume to free energy in the same range as that found in protein stability studies. Thus, similar free energy penalties results from creating interior cavities both for protein folding and ligand binding. For the cavities in the S1' pocket, the volume change of the cavity is the largest for KNI-272, which has the largest DA(S1'), and the smallest for Ro 31-8959, which has the smallest DA(S1') and the largest P' substituent. The interior cavities can also be predicted based on the wildtype structures; in the S1' binding pocket, the predicted cavities are larger than the cavities determined from the mutant structures, in which the structures have been allowed to adjust to mutation. This means that there is significant relaxation in this region. In the S1 binding pocket, there is not a large difference between the predicted and the actual mutant cavity sizes. Both of these problems - the relative stability of two structures and the difference in the binding affinity between two different mutants - involves calculating a free energy difference. The calculation of free energy involves specialized simulation techniques which require more computer processing time than typical simulations but also provide an additional level of information.

Benefits of Scalable Increases in Compute Power

The calculations of the free energy changes described here require large amounts of computer time in order to adequately sample over all the degrees-of-freedom of the system. The number of degrees-of-freedom can be very large. The machines used for these calculations, running at about 100 megaflops, take from three weeks (CPU time) for the mutation calculations t three months (CPU time) for the conformational change calculation. In order to make these calculations routine enough to be used as part of a drug design process, machine speeds will need to be increased by two orders of magnitude to ten gigaflops.